Rectifiability of Optimal Transportation Plans*

ROBERT J. McCANN,† BRENDAN PASS‡AND MICAH WARREN§

March 25, 2010

Abstract

The purpose of this note is to show that the solution to the Kantorovich optimal transportation problem is supported on a Lipschitz manifold, provided the cost is C^2 with non-singular mixed second derivative. We use this result to provide a simple proof that solutions to Monge's optimal transportation problem satisfy a change of variables equation almost everywhere.

1 Introduction

Given Borel probability measures μ^+ and μ^- on smooth n-dimensional manifolds M^+ and M^- respectively and a cost function $c : M^+ \times M^- \to \mathbf{R}$, the Kantorovich problem is to pair the two measures as efficiently as possible relative to c. A precise formulation requires some notation. For a measure γ on $M^+ \times M^-$, we define the marginals of γ to be its push forwards under the canonical projections π^+ and π^-; put another way, the

marginals are measures on M^+ and M^- respectively given by the formulae $\pi_\#^+\gamma(A) = \gamma(A \times M^-)$ and $\pi_\#^-\gamma(B) = \gamma(M^+ \times B)$ for all Borel sets $A \subset M^+, B \subset M^-$. The Kantorovich problem is then to minimize the functional

$$\int_{M^+ \times M^-} c(x,y)d\gamma(x,y) \tag{1}$$

among all measures γ on $M^+ \times M^-$ whose marginals are $\pi_\#^+\gamma = \mu^+$ and $\pi_\#^-\gamma = \mu^-$.

Under fairly weak conditions, it is straightforward to show that a solution to this problem exists. In this paper, we study what can be said about that solution under a certain non-degeneracy condition on the cost function, which was originally introduced in an economic context by McAfee and McMillan [25] and later rediscovered by Ma, Trudinger and Wang [26]; in the terminology of Ma, Trudinger and Wang, it is also known as the (A2) condition.

In what follows, $D^2_{xy}c(x_0, y_0)$ will denote the n by n matrix of mixed second order partial derivatives of the function c at the point $(x_0, y_0) \in M^+ \times M^-$; its (i,j)th entry is $\frac{d^2 c}{dx^i dy^j}(x_0, y_0)$.

Definition 1.1. *Assume $c \in C^2(M^+ \times M^-)$. We say that c is non-degenerate at a point $(x_0, y_0) \in M^+ \times M^-$ if $D^2_{xy}c(x_0, y_0)$ is nonsingular; that is if $det(D^2_{xy}c(x_0, y_0)) \neq 0$.*

For a probability measure γ on $M^+ \times M^-$ we will denote by spt(γ) the support of γ; that is, the smallest closed set $S \subseteq M^+ \times M^-$ such that $\gamma(S) = 1$.

Our main result is:

Theorem 1.2. *Suppose $c \in C^2(M^+ \times M^-)$ and μ^+ and μ^- are compactly supported; let γ be a solution of the Kantorovich problem. Suppose $(x_0, y_0) \in$ spt(γ) and c is non-degenerate at (x_0, y_0). Then there is a neighbourhood N of (x_0, y_0) such that $N \cap$ spt(γ) is contained in an n-dimensional Lipschitz submanifold. In particular, if $D^2_{xy}c$ is nonsingular everywhere, sptγ is contained in an n-dimensional Lipschitz submanifold.*

The proof of this theorem is based on an idea of Minty [30], which was also used by Alberti and Ambrosio to show that the graph of any monotone function $T : \mathbf{R}^n \to \mathbf{R}^n$ is contained in a Lipschitz graph over the diagonal $\Delta = \{u = \frac{x+y}{\sqrt{2}} : (x,y) \in \mathbf{R}^n \times \mathbf{R}^n\}$ [4].

The non-degeneracy condition can be viewed as a linearized version of the twist condition, which asserts that the mapping $y \in M^- \longmapsto D_x c(x, y)$ is injective. Under suitable regularity conditions on the marginals, Levin [21] showed that the twist condition ensures that the solution to the Kantorovich problem is concentrated on the graph of a function and is therefore unique; see also Gangbo [17]. For the past two decades, the regularity of these maps has been an active area of investigation. Regularity results were proven for the quadratic cost function by Caffarelli [9][10][11], Delanoë [14][15] and Urbas [36] and for another special cost function by Wang [37]. These were then generalized by Ma, Trudinger and Wang [26], who discovered a fourth order differential condition on the cost function that ensures the optimal map is smooth, provided the marginals are sufficiently regular [26][34]. Our results assert that something can be said about the smoothness of the support even without these strong conditions on the cost and the marginals, provided that one is willing to view the support as a submanifold rather than a graph.

In one dimension, non-degeneracy implies twistedness, as was noted by many authors, including Spence [33] and Mirrlees [31], in the economics literature; see also [29]. In higher dimensions, this is no longer true; the non-degeneracy condition will imply that the map $y \in M^- \longmapsto D_x c(x, y)$ is locally injective but not necessarily globally. Non-degeneracy was a hypothesis in the smoothness proof in [26], but does not seem to have received much attention in higher dimensions before then. While our result demonstrates that the non-degeneracy condition is enough to ensure that solutions still have certain regularity properties, we will show by example that the uniqueness result that follows from twistedness can fail for non-degenerate costs which are not twisted. The twist condition is asymmetric in x and y; that is, there are cost functions for which the map $y \in M^- \longmapsto D_y c(x, y)$ is injective but $x \in M^+ \longmapsto D_y c(x, y)$ is not. However, since $(D^2_{xy} c)^T = D^2_{yx} c$ the non-degeneracy condition is certainly symmetric in x and y. In view of this, it is not surprising that the twist condition can only be used to show solutions are concentrated on the graphs of functions of y over x whereas the non-degeneracy condition implies solution are concentrated on n-dimensional submanifolds, a result that does not favour either variable over the other.

Smooth optimal maps solve certain Monge-Ampère type equations. Typically, an optimal map will be differentiable almost everywhere, but may not be smooth. It has proven useful to know when non-smooth optimal maps solve the corresponding equations almost everywhere. Formally, the link between optimal transportation and these equations was observed by Brenier

[8], then Gangbo and McCann [18], and they were studied in detail by Ma, Trudinger and Wang [26]. An important step in showing that an optimal map solves a Monge-Ampère type equation is first showing that it solves the Jacobian — or change of variables — equation. An injective Lipschitz function satisfies the change of variables formula almost everywhere, so some sort of Lipschitz rectifiability for the graphs of optimal maps is a useful tool in resolving this question. As an application of Theorem 1.2, we provide a simple proof that optimal maps satisfy the change of variables formula almost everywhere.

This work is related to another interesting line of research. A measure γ on the product $M^+ \times M^-$ is called *simplicial* if it is extremal among the convex set of all measures which share its marginals. There are a number of results describing simplicial measures and their supports [16][22][7][20][3]. One consequence is that the support of simplicial measures are in some sense small; in particular, the support of a simplicial measure on $[0, 1] \times [0, 1]$ must have two-dimensional Lebesgue measure zero [22][20]. However, any measure supported on the graph of a function is simplicial and it is known that there exist functions whose graphs have Hausdorff measure $2 - \epsilon$, for any $\epsilon > 0$ [1]. For any cost, the Kantorovich functional is linear and is hence minimized by some simplicial measure. Conversely, any simplicial measure is the solution to a Kantorovich problem for some continuous cost function, and so by the remarks above there are continuous cost functions whose optimizers are supported on sets of Hausdorff dimension $2 - \epsilon$. On the other hand, an immediate consequence of our result is that the support of optimizers of Kantorovich problems with non-degenerate C^2 costs have Hausdorff dimension at most n.

The result of Ma, Trudinger and Wang proving smoothness of the optimal map under certain conditions immediately implies that the support of the optimizer has Hausdorff dimension n; however, the proof of this result requires that the marginals be C^2 smooth. Under the same assumptions on the cost functions but weaker regularity conditions on the marginals, Loeper [24] and Liu [23] have demonstrated that the optimal map is Hölder continuous for some Hölder constant $0 < \alpha < 1$. It is worth noting that there are examples of functions on $\mathbf{R}^n$ [1] which are Hölder continuous with exponent α but whose graphs have Hausdorff dimension $n + 1 - \alpha$, so the latter results do not imply that the Hausdorff dimension of the optimizer must be n.

In the second section of this manuscript we prove Theorem 1.2 while Section 3 is devoted to discussion and examples. In the final section, we use

Theorem 1.2 to provide a simple proof that optimal maps satisfy a prescribed Jacobian equation almost everywhere.

We are pleased to acknowledge that our interest in this topic was stimulated in part by a fruitful discussion between one of the authors and Ivar Ekeland.

2 Lipschitz Rectifiability of Optimal Transportation Plans

We now prove Theorem 1.2. Note that γ minimizes the Kantorovich functional if and only if it maximizes the corresponding functional for $b(x, y) = -c(x, y)$. To simplify the computation, we consider γ that maximizes b.

Our proof relies on the *b-monotonicity* of the supports of optimal measures:

Definition 2.1. *A subset S of $M^+ \times M^-$ is b-monotone if all $(x_0, y_0), (x_1, y_1) \in S$ satisfy $b(x_0, y_0) + b(x_1, y_1) \geq b(x_0, y_1) + b(x_1, y_0)$.*

It is well known that the support of any optimizer is b-monotone [32], provided that the cost is continuous and the marginals are compactly supported. The reason for this is intuitively clear; if $b(x_0, y_0) + b(x_1, y_1) > b(x_0, y_1) + b(x_1, y_0)$ then we could move some mass from (x_0, y_0) and (x_1, y_1) to (x_0, y_1) and (x_1, y_0) without changing the marginals of γ and thus increase the integral of b.

The strategy of our proof is to change coordinates so that locally $b(x, y) = x \cdot y$, modulo a small perturbation. We then switch to diagonal coordinates $u = x + y, v = x - y$ and show that the monotonicity condition becomes a Lipschitz condition for v as a function of u. This trick dates back to Minty who used it to study monotone operators on Hilbert spaces [30]; more recently, Alberti and Ambrosio used it to investigate the fine properties of monotone functions on $\mathbf{R}^n$ [4].

We are now ready to prove Theorem 1.2:

Proof. Choose (x_0, y_0) in the support of γ. Changing coordinates in a neighbourhood of y_0 yields $D^2_{xy} b(x_0, y_0) = I$ without loss of generality. We then have $b(x, y) = x \cdot y + G(x, y)$, where $D^2_{xy} G \to 0$ as $(x, y) \to (x_0, y_0)$. Set $u\sqrt{2} = x + y$ and $v\sqrt{2} = y - x$. Given $\epsilon > 0$, choose a convex neighbourhood N of (x_0, y_0) such that $||D^2_{xy} G|| \leq \epsilon$ on N. We will show that $\gamma \cap N$

5

is contained in a Lipschitz graph of v over u; hence, u and v serve as local coordinates for our submanifold. Take (x, y) and $(x', y') \in N \cap spt\gamma$. Then, by b-monotonicity, we have $b(x, y) + b(x', y') \geq b(x, y') + b(x', y)$, hence

$$
\begin{aligned}
x \cdot y &+ G(x, y) + x' \cdot y' + G(x', y') \\
&\geq x \cdot y' + G(x, y') + x' \cdot y + G(x', y).
\end{aligned}
$$

Setting $\Delta x = x' - x$, $\Delta y = y' - y$, $\Delta u = u' - u$, $\Delta v = v' - v$, and rewriting yields

$$
(\Delta x) \cdot (\Delta y) + (\Delta x) \cdot \int_0^1 \int_0^1 D_{xy}^2 G[x + s\Delta x, y + t\Delta y](\Delta y) ds dt \geq 0 \qquad (2)
$$

which simplifies to: $\Delta x \cdot \Delta y \geq -\epsilon |\Delta x||\Delta y|$.

Observe that $\Delta y \sqrt{2} = \Delta u + \Delta v$ and $\Delta x \sqrt{2} = \Delta u - \Delta v$. Now,

$$
\begin{aligned}
|\Delta u|^2 - |\Delta v|^2 &= 2(\Delta x) \cdot (\Delta y) \\
&\geq -2\epsilon |\Delta x||\Delta y| \\
&= -\epsilon |\Delta u - \Delta v||\Delta u + \Delta v| \\
&\geq -\epsilon [|\Delta u|^2 + |\Delta v|^2]
\end{aligned}
$$

The last inequality follows by squaring the absolute values of each side and expanding the first term. Rearranging yields $(1 + \epsilon)|\Delta u|^2 \geq (1 - \epsilon)|\Delta v|^2$, the desired result.

Note that v may not be everywhere defined; that is, for certain values of u there may be no corresponding v in $spt(\gamma)$. However, the function $v(u)$ can be extended by Kirzbraun's theorem and hence we can conclude that $spt(\gamma)$ is contained in the graph of a Lipschitz function of v over u. □

Remark 2.1. *Note that the only property of optimal transportation plans used in the proof is b-monotonicity, so we have actually proven that any b-monotone subset of $M^+ \times M^-$ is contained in an n-dimensional Lipschitz submanifold, provided b is non-degenerate.*

3 Discussion and Examples

For twisted costs, one can show that $spt(\gamma)$ is concentrated on the graph of a function, provided the marginal μ^+ does not charge sets whose dimension

is less than or equal to $n-1$[17] [21] [26] [3] [27] [18][1]; however, this can fail if μ^+ charges small sets. On the other hand, notice that our proof did not require any regularity hypotheses on the marginals.

In the example below, we exhibit a non-degenerate cost which is not twisted. We use this example to illustrate how, in this setting, solutions may be supported on submanifolds which are are not necessarily graphs. In addition, we show that these solutions may not be unique. We can view this example as expressing an optimal transportation problem on a right circular cylinder via its universal cover, which is $\mathbf{R}^2$. The non-twistedness of the cost and non-uniqueness of the solution arise because different points in the universal cover correspond to the same point in the cylinder and are therefore indistinguishable by our cost function. In fact, if we expressed the problem on the cylinder, we would have a twisted cost function and therefore a unique solution.

Example 3.1. *Let* $M^\pm = \mathbf{R}^2$ *and* $c(x,y) = e^{x_1+y_1}\cos(x_2-y_2) + \frac{e^{2x_1}}{2} + \frac{e^{2y_1}}{2}$. *Then* $D_x c(x,y) = (e^{x_1+y_1}\cos(x_2-y_2) + e^{2x_1}, -e^{x_1+y_1}\sin(x_2-y_2))$, *so* $y \in M^- \longmapsto D_x c(x,y)$ *is not injective and* c *is not twisted. However, note that* $D^2_{xy}c(x,y) =$

$$\begin{bmatrix} e^{x_1+y_1}\cos(x_2-y_2) & e^{x_1+y_1}\sin(x_2-y_2) \\ -e^{x_1+y_1}\sin(x_2-y_2) & e^{x_1+y_1}\cos(x_2-y_2) \end{bmatrix}$$

Therefore, $\det D^2_{xy}c(x,y) = e^{2(x_1+y_1)} > 0$ *for all* (x,y), *so* c *is non-degenerate. Optimal measures for* c, *then, must be supported on* 2-*dimensional Lipschitz submanifolds, but we will now exhibit an optimal measure whose support is not contained in the graph of a function.*

Now let M *be the union of the three graphs:*

$$G_1 : y_1 = x_1, y_2 = x_2 + \pi \tag{3}$$
$$G_2 : y_1 = x_1, y_2 = x_2 + 3\pi \tag{4}$$
$$G_3 : y_1 = x_1, y_2 = x_2 + 5\pi \tag{5}$$

Clearly, M *is a smooth* 2-*d submanifold but not a graph. However,* $c(x,y) \geq -e^{x_1+y_1} + \frac{e^{2x_1}}{2} + \frac{e^{2y_1}}{2} \geq \frac{(e^{x_1}-e^{y_1})^2}{2}$ *and we have equality on* M. *Therefore, any probability measure whose support is concentrated on* M *is optimal for its marginals.*

---

[1]In fact, this condition on the regularity of μ^+ has recently been sharpened [19].

We now show that optimal measures supported on M may not be unique. Let $S = \{((x_1, x_2), (y_1, y_2)) | 0 \le x_1 \le 1, 0 \le x_2 \le 4\pi\}$. Note that

$$M \cap S = (G_1 \cap S) \cup (G_2 \cap S) \cup (G_3 \cap S).$$

consists of 3, flat 2-d regions. Let γ be uniform measure on these regions. Now, let $\overline{\gamma_1}$ be uniform measure on the the first half of $G_1 \cap S$; that is, on

$$G_1 \cap \{((x_1, x_2), (y_1, y_2)) | 0 \le x_1 \le 1, 0 \le x_2 \le 2\pi\}.$$

Let $\overline{\gamma_3}$ be uniform measure on the the second half of $G_3 \cap S$, or

$$G_3 \cap \{((x_1, x_2), (y_1, y_2)) | 0 \le x_1 \le 1, 2\pi \le x_2 \le 4\pi\}.$$

Take $\overline{\gamma_2}$ to be twice uniform measure on $G_2 \cap S$ and set $\overline{\gamma} = \overline{\gamma_1} + \overline{\gamma_2} + \overline{\gamma_3}$. Then γ and $\overline{\gamma}$ share the same marginals and are both optimal measures. Furthermore, any convex combination $t\gamma + (1-t)\overline{\gamma}$ will also share the same marginals and will be optimal as well.

The next example is similar in that the cost function is non-degenerate but not twisted. However, this cost *would* be twisted if we exchanged the roles of x and y. This demonstrates that, unlike non-degeneracy, the twist condition is not symmetric in x and y. For this cost function, solutions will be unique as long as the *second* marginal does not charge small sets.

Example 3.2. *Let $M^{\pm} = \mathbf{R}^2$ and $c(x, y) = -(x_1 \cos(y_1) + x_2 \sin(y_1))e^{y_2} + \frac{e^{2y_2}}{2} + \frac{x_1^2 + x_2^2}{2}$. Note that $\det D_{xy}^2 c(x, y) = -e^{2y_2} < 0$, so c is non-degenerate. However, $D_x c(x, y) = (-\cos(y_1)e^{y_2} + x_1, -\sin(y_1)e^{y_2} + x_2)$, so $y \in M^{-} \longmapsto D_x c(x, y)$ is not injective and c is not twisted. On the other hand, $D_y c(x, y) = ((x_1 \sin(y_1) + x_2 \cos(y_1))e^{y_2}, -(x_1 \cos(y_1) + x_2 \sin(y_1))e^{y_2} + e^{2y_2})$ and so $x \in M^{+} \longmapsto D_y c(x, y)$ is injective. This implies that solutions are supported on graphs of x over y but that these graphs are not necessarily invertible. In fact, $c(x, y) \ge \frac{((x_1^2 + x_2^2)^{\frac{1}{2}} - e^{y_2})^2}{2} \ge 0$, where equality holds if and only if $\cos(y_1) = \frac{x_1}{(x_1^2 + x_2^2)^{\frac{1}{2}}}$, $\sin(y_1) = \frac{x_2}{(x_1^2 + x_2^2)^{\frac{1}{2}}}$, and $(x_1^2 + x_2^2)^{\frac{1}{2}} = e^{y_2}$. This set of equality is a non-invertible graph of x over y; any measure whose support is contained in this graph is optimal for its marginals. Note that as any minimizer for this problem must be supported on this graph, the solution is unique [3].*

Remark 3.3. *For twisted costs with regular marginals, any solution is concentrated on the graph of a particular function [26]. It is not hard to show that at most one measure with prescribed marginals can be supported on such a graph; hence, uniqueness of the optimizer follows immediately.*

While our result asserts that for non-degenerate costs the solution concentrates on some n-dimensional Lipschitz submanifold, the proof says little more about the submanifold itself. In contrast to the twisted setting, then, our result cannot be used to deduce a uniqueness argument. Furthermore, as the example above shows, even if we do know the support of the optimizer explicitly, solutions may not be unique if this support is not concentrated on the graph of a function.

Theorem 1.2 also says something about problems where $D^2_{xy}c$ is allowed to be singular, but where the gradient of its determinant is non-zero at the singular points. In this case, the implicit function theorem implies that the set where $D^2_{xy}c$ is singular has Hausdorff dimension $2n - 1$. Theorem 1.2 is valid wherever $D^2_{xy}c$ is nonsingular, so that the optimal measure is concentrated on the union of a smooth $2n - 1$ dimensional set and an n dimensional Lipschitz submanifold. For example, when $n = 1$, this shows that the support of the optimal measure is 1 dimensional.

4 A Jacobian equation

We now provide a simple proof that an optimal map satisfies a prescribed Jacobian equation almost everywhere. This result was originally proven for the quadratic cost in $\mathbf{R}^n$ by McCann [28], and for the quadratic cost on a Riemannian manifold by Cordero-Erasquin, McCann and Schmuckenschläger [13]. Cordero-Erasquin generalized this approach to deal with strictly convex costs on $\mathbf{R}^n$ [12]; see also [2]. It was observed by Ambrosio, Gigli and Savare that this can be deduced from results in [5] and [6] when the optimal map is approximately differentiable, which is true even for some non-smooth costs. Our method works only when the cost is C^2 and non-degenerate, but has the advantage of a simpler proof, relying only on the area/coarea formula for Lipschitz functions.

For a Jacobian equation to make sense, the solution must be concentrated on the graph of a function, and that function must be differentiable in some sense, at least almost everywhere. A twisted cost suffices to ensure the first condition. The second follows from the smoothness and non-degeneracy of

c. Recall that for a twisted cost the optimal map has the form $T(x) = c\text{-}exp_x(Du(x))$; as $c\text{-}exp_x(\cdot)$ is the inverse of $y \longmapsto D_x c(x,y)$, its differentabiliy follows from the non-degeneracy of c and the inverse function theorem. The almost everywhere differentiability of $Du(x)$ (or, equivalently, the almost everywhere twice differentiability of u) follows from C^2 smoothness of c; u takes the form $u(x) = \inf_y(c(x,y) - v(y))$ for some function $v(y)$ and is hence semi-concave [18]. In the present context, we need only the weaker condition that the optimal map is continuous almost everywhere; its differentiability will follow from Theorem 1.2.

Proposition 4.1. *Assume that the cost is non-degenerate and that an optimizer γ is supported on the graph of some function $T : dom(T) \to M^-$ which is injective and continuous when restricted to a set $dom(T) \subseteq M^+$ of full Lebesgue measure. Suppose that the marginals are absolutely continuous with respect to volume; set $d\mu^+ = f^+(x)dx$ and $d\mu^- = f^-(y)dy$. Then, for almost every x, $f^+(x) = |detDT(x)|f^-(T(x))$.*

Proof. Choose a point x where T is continuous and a neighbourhood U^- of $T(x)$ such that for $U^+ = T^{-1}(U^-)$, the part of the optimal graph contained in $U^+ \times U^-$ lies in a Lipschitz graph $v = G(u)$ over the diagonal $\Delta = \{u = \frac{x+y}{\sqrt{2}} : (x,y) \in U^+ \times U^-\}$, after a change of coordinates. Now $x = \frac{u+v}{\sqrt{2}}$ and $y = \frac{u-v}{\sqrt{2}}$, so the optimal measure is supported on the graph of the Lipschitz function $(x,y) = (F^+(u), F^-(u)) := (\frac{u+G(u)}{\sqrt{2}}, \frac{u-G(u)}{\sqrt{2}})$. By projecting onto the diagonal, we obtain a measure ν on Δ that pushes forward to $\mu^+|_{U^+}$ and $\mu^-|_{U^-}$ under the Lipschitz mappings F^+ and F^-, respectively. Now, as F^+ is Lipschitz, the image of any zero volume set must also have zero volume; as $\mu^+|_{U^+}$ is absolutely continuous with respect to Lebesgue, ν must be as well; we will write $\nu = h(u)du$. Now, for almost every $x \in U^+$ there is a unique $y = T(x)$ such that $(x,y) \in \text{spt}(\gamma)$ and hence a unique $u = \frac{x+y}{\sqrt{2}}$ on the diagonal such that $x = F^+(u)$. It follows that the map F^+ is one to one almost everywhere and so for every set $A \subseteq \Delta$ we have $\int_A h(u)du = \int_{F^+(A)} f^+(x)dx$. But the right hand side is $\int_A f^+(F^+(u))|detDF^+(u)|du$ by the area formula; as A was arbitrary, this means $h(u) = f^+(F^+(u))|detDF^+(u)|$ almost everywhere. Similarly, $h(u) = f^-(F^-(u))|detDF^-(u)|$ almost everywhere, hence

$$f^+(F^+(u))|detDF^+(u)| = f^-(F^-(u))|detDF^-(u)|$$

almost everywhere. As the image under F^+ of a negligible set must itself be

negligible, we have

$$f^+(x)|\det DF^+((F^+)^{-1}(x))| = f^-(F^-((F^+)^{-1}(x)))|\det DF^-((F^+)^{-1}(x))|$$
(6)

for almost all x. Note that as F^+ is one to one almost everywhere and $F^+(\{u \in \Delta : \det DF^+(u) = 0\})$ has measure zero by the area formula, $(F^+)^{-1}$ is differentiable almost everywhere. As $T \circ F^+ = F^-$, it follows that T is differentiable almost everywhere and

$$\det DT(F^+(u))\det DF^+(u) = \det DF^-(u)$$

whenever F^+ and F^- are differentiable at u and T is differentiable at $F^+(u)$. Hence,

$$\det DT(x)\det DF^+((F^+)^{-1}(x)) = \det DF^-((F^+)^{-1}(x)) \qquad (7)$$

for all x such that T is differentiable at x and F^+ and F^- are differentiable at $(F^+)^{-1}(x)$. T is differentiable for almost every x, F^+ and F^- are differentiable for almost every u and F^+ is Lipschitz; it follows that the above holds almost everywhere. Now, combining (6) and (7) we obtain $f^+(x) = |\det DT(x)|f^-(T(x))$ for almost every x. $\qquad \square$

Remark 4.1. *Note that the preceding proposition does not require that continuity of T extend outside dom(T). Thus it applies to $T = Du$, for example, where u is an arbitrary convex function and dom(T) is its domain of differentiability.*

References

[1] Adler, R.J. *The geometry of random fields.* Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Ltd., Chichester, 1981.

[2] Agueh, M. Existence of solutions to degenerate parabolic equation via the Monge-Kantorovich theory. *PhD Dissertation,* Georgia Institute of Technology, (2002).

[3] Ahmad, N. Kim, H.K., and McCann, R.J., Extremal doubly stochastic measures and optimal transportation. Preprint.

[4] Alberti, G. and Ambrosio, L., A geometrical approach to monotone functions on $\mathbf{R}^n$. *Math. Z.*, **230** (1999), 259-316.

[5] Ambrosio, L. Fusco, N., and Pallara, D. *Functions of bounded variation and free discontinuity problems.* Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2000.

[6] Ambrosio, L., Gigli, N., and Savaré, G. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Z̈urich. Birkhäuser Verlag, Basel, 2005.

[7] Beneš, V. and Štěpàn, J. The support of extremal probability measures with given marginals. In *Mathematical Statistics and Probability Theory,* A (eds. M.L. Puri et. al.), (1987), 33-41.

[8] Brenier, Y. Decomposition polaire et rearrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Pair. Ser. I Math.*, **305** (1987), 805-808.

[9] Caffarelli, L.A. The regularity of mappings with a convex potential. *J. Amer. Math. Soc*, **5** (1992), 99-104.

[10] Caffarelli, L.A. Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.,* **45** (1992), 1141-1151.

[11] Caffarelli, L.A. Boundary regularity of maps with convex potentials-II. *Ann. of Math. (2)*, **144** (1996), 453-496.

[12] Cordero-Erausquin, D. Non-smooth differential properties of optimal transport. In *Recent Advances in the Theory and Application of Mass Transport,* vol 353 of *Contemp. Math.,* Amer. Math. Soc., Providence, RI, 2004, 61-71.

[13] Cordero-Erausquin, D., McCann, R. J., and Schmuckenschläger, M. A Riemannian interpolation inequality a la Borell, Brascamp and Lieb. *Invent. Math.*, **146**, 2 (2001), 219-257.

[14] Delanoë, P. Classical solvability in dimension two of the second boundary-value problem associated with the Monge-Ampere operator. *Ann. Inst. H. Poincare Anal. Non Lineaire*, **8** (1991), 442-457.

[15] Delanoë, P. Gradient rearrangment for diffeomorphisms of a compact manifold. *Differential Geom. Appl.*, **20** (2004), 145-165.

[16] Douglas, R.D. On extremal measures and subspace density. *Michigan Math. J.*, **11** (1964), 243-246.

[17] Gangbo, W. *Habilitation Thesis*, Universite de Metz, (1995).

[18] Gangbo, W. and McCann, R.J. The geometry of optimal transportation. *Acta Math.*, **177** (1996), 113-161.

[19] Gigli, N. On the inverse implication of Brenier-McCann theorems and the structure of $(\mathcal{P}_2(M), W_2)$. Preprint.

[20] Hestir, K. and Williams, S.C. Supports of doubly stochastic measures. *Bernoulli*, **1** (1995) 217-243.

[21] Levin, V. Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem. *Set-Valued Analysis*, **7**, 1 (1999), 7-32.

[22] Lindenstrauss, J. A remark on extreme doubly stochastic measures. *Amer. Math. Monthly*, **72** (1965), 379-382.

[23] Liu, J. Hölder regularity in optimal mappings in optimal transportation. To appear in *Calc. Var. Partial Differential Equations.*

[24] Loeper, G. On the regularity of maps solutions of optimal transportation problems. *Acta Math.*, **202** (2009), 241-283.

[25] McAfee, R.P. and McMillan, J. Multidimensional incentive compatibility and mechanism design. *J. Econom. Theory*, **46** (1988), 335-354.

[26] Ma, X-N., Trudinger, N., and Wang, X-J. Regularity of potential functions of the optimal transportation problem. *Arch. Rational Mech. Anal.*, **177** (2005), 151-183.

[27] McCann, R.J. Existence and uniqueness of monotone measure preserving maps. *Duke Math. J.* **80** (1995), 309-323.

[28] McCann, R.J. A convexity theory for interacting gases and equilibrium crystals. *Adv. Math.*, **128**, 1 (1997), 153-179.

[29] McCann, R.J. Exact solutions to the transportation problem on the line. *Proc. Royal Soc. London Ser. A* **455** (1999), 1341-1380.

[30] Minty, G. J. Monotone (nonlinear) operators in Hilbert space. *Duke Math. J.*, **29** (1962), 341-346.

[31] Mirrlees, J.A. An exploration of optimum income taxation, *Rev. Econom. Stud.*, **38** (1971), 175-208.

[32] Smith, C. and Knott, M. On Hoeffding-Frechet bounds and cyclic monotone relations. *J. Multivariate Anal.*, **40** (1992), 328-334.

[33] Spence, M. Competitive and optimal responses to signals: An analysis of efficiency and distribution, *J. Econom. Theory*, **7** (1974), 296-332.

[34] Trudinger, N., and Wang, X-J. On the second boundary value problem for Monge-Ampere type equations and optimal transportation. To appear in *Ann.Sc. Norm. Super. Pisa Cl. Sci.*

[35] Trudinger, N., and Wang, X-J. On strict convexity and C^1-regularity of potential functions in optimal transportation. To appear in *Arch. Rational Mech. Anal.*

[36] Urbas, J. On the second boundary value problem for equations of Monge-Ampere type. *J. Reine Angew. Math.*, **487** (1997), 115-124.

[37] Wang, X-J. On the design of a reflector antenna. *Inverse problems*, **12**, 3 (1996), 351-375.